

الفصل السابع

المقارنة بين اللغة الطبيعية واللغة المحكومة في تنظيم

واسترجاع المعلومات

- أولاً : نبذه عن تاريخ نظم استرجاع المعلومات.
- ثانياً : استرجاع المعلومات والمنطق البوليني.
- ثالثاً : أنواع الاسترجاع.
- رابعاً : تقسيمات الشبكات وشبكة الانترنت.
- خامساً : لغة التكشيف.
- سادساً : الحاجة إلى المصطلحات المحكومة.
- سابعاً : توليد المصطلحات وزيادة كفاءة التعبير والنمو.
- ثامناً : بعض جوانب المقارنة بين اللغة الطبيعية والمصطلحات المحكومة.
- تاسعاً : تحسين الاسترجاع عن طريق تحسين عمليات ولغات الأسئلة وتحسين عمليات ولغات النص.

مراجع الفصل

obbeikandi.com

الفصل السابع

المقارنة بين اللغة الطبيعية واللغة المحكومة

فى نظم استرجاع المعلومات (*)

يعتبر التوثيق واسترجاع المعلومات هما أساس علم المعلومات كما جاء فى مقالات كل من جيسى شيرا J. Shera وفارسمان جلين J. Wurman والتي أيدهما حشمت قاسم (١٩٩٠) فى كتابه مدخل لدراسة المكتبات وعلم المعلومات (ص ٨٣)، وإذا كان كاتب هذه السطور قد سجل ذلك عن التوثيق فى تاريخ علم المعلومات فى الفصل الأول، فهو هنا يتناول نظم استرجاع المعلومات وأساليب التكشيف بما يتضمنه من مقارنة اللغة الطبيعية بالمصطلحات المحكومة.

أولاً : نبذة عن تاريخ نظم استرجاع المعلومات :

تعود الجذور الأولى لاسترجاع المعلومات إلى الدراسات المقارنة بين التصنيف والتكشيف فى أوائل الخمسينيات عندما قام مورتيمرتأوبه Mortimer Taube بمقارنة نظام تكشيف المصطلح الواحد Uniterm Indexing System بالتصنيف التقليدى وبرؤوس الموضوعات والتي تمت فى الولايات المتحدة، أما فى بريطانيا فقد قام كرانفيلد Cranfield بمقارنة المصطلحات الواحدة Uniterms بالتصنيف العشرى العالمى (UDC)، كما تعود هذه الجذور أيضاً إلى المدخل الإحصائى للباحث هانزبيتر لوهن Hans Peter Luhn وأكتشف العلاقة بين توزيعات كلمات الوثيقة للواصفات Descriptors اللازمة لوصف وتحليل الوثيقة، ولابد أخيراً من الإشارة إلى أن تطور تكنولوجيا المعلومات والاتصالات كان لها ومازال قوة دافعة فى تحسين استرجاع

(*) هذه الدراسة موجز لبحث الكاتب المنشور بمجلة مكتبة الملك فهد الوطنية مجلد ٣ عدد ١.

المعلومات منذ الستينيات حتى الآن أى منذ نموذج المضاهاة التقليدية حتى النظم الضخمة التى تعتمد على الوب.

كما يمكن اعتبار عملية إسترجاع المعلومات كحالة خاصة من سلوك البحث عن المعلومات حيث تستخدم الحاسبات فى البحث والاسترجاع والاتاحة للمعلومات المطلوبة، ونظام الإسترجاع يقوم بتطبيق الجورزم - هو أسلوب مضاهاة Matching- للمقارنة بين الحاجة المعلوماتية والتمثيلات الوثائقية، ونتيجة ذلك مجموعة من الوثائق التى تضاهى السؤال، وتتم تلك الاتاحة إما عن طريق لغة الأسئلة المعتمدة على المنطق البوليني Boolean Logic أو أن تتم عن طريق نماذج الإسترجاع models التى تؤدى إلى قائمة تنازلية عن طريق صلاحية relevance هذه الوثائق عن غيرها، وإن كانت نماذج الإسترجاع التى تؤدى إلى القائمة التنازلية حسب الصلاحية تجريبية فى المختبر أكثر منها صالحة لامكان التطبيق على مستوى واسع ذلك لأنها تعتمد على قياسات الأداء لكل من الاستدعاء والدقة وهو كما يلى:

عدد الوثائق الصالحة المسترجعة

$$\frac{\text{عدد الوثائق الصالحة المسترجعة}}{\text{العدد الكلى للوثائق المسترجعة فى المجموعة}} = \text{Recall الاستدعاء}$$

العدد الكلى للوثائق المسترجعة فى المجموعة

عدد الوثائق الصالحة المسترجعة

$$\frac{\text{عدد الوثائق الصالحة المسترجعة}}{\text{العدد الكلى للوثائق المسترجعة}} = \text{الدقة Precision}$$

العدد الكلى للوثائق المسترجعة

ثانياً : إسترجاع المعلومات والمنطق البوليني:

تعمل معظم نظم إسترجاع المعلومات على مبدأ مضاهاة النص text matching حيث يستخدم مصطلح بحثى (من كلمة أو كلمتين أو ثلاثة كلمات لتكوين عبارة) ويشكل ذلك مدخلات، على أن يقوم نظام الإسترجاع بإعداد استجابة تشمل مجموعة من التسجيلات من قاعدة البيانات (وهذه يمكن أن تكون وثائق أو مراجع للوثائق) والتى تجيب على المصطلح موضع السؤال.

فالاسترجاع الناجح عن طريق المضاهاة matching باستخدام الأدوات التالية:

الأدوات	نوع البحث	رسم فين Venn	المعنى
و AND	الاقتران	(أ) (ب) 	النتائج المنطقية: ويتمثل بالرموز أ و ب / أ، ب / أ × ب / (أ) (ب) وهنالك لا بد أن يكون المصطلحان أ و ب موجودين بالوثيقة للمضاهاة.
أو OR	إضافة	المرأة الانثى 	المجموع المنطقية: ويتمثل بالرموز أ أو ب / أ + ب وفى هذه الحالة يتطلب الأمر أن يكون واحد فقط من مصطلحي التكشيف أ أو ب موجوداً بالوثيقة لتنتم المضاهاة.
وليس NOT	طرح	الاعلان الترويج 	الاستبعاد المنطقية: ويتمثل بالرموز أ وليس ب / أ - ب وفى هذه الحالة يجب تخصيص المصطلح أ، وأن يتم هذا التخصيص فى غياب المصطلح ب وذلك لإمكانية المضاهاة.

أدوات المنطق البوليني Boolean Logic

هذا ويلاحظ أن كل برنامج جاهز (أو تصنيف) له قواعد أولوياته الخاصة (وعلى سبيل المثال فإن أداة And يمكن أن يتم استخدامها قبل or ويعتمد البحث الناجح على مراعاة هذه القواعد والإفادة المناسبة من الأقواس، وذلك لأن الأقواس تقدم غالباً تحديداً للأولية من منظور الباحث.

أما الرابط الثالث وهو Not فيعنى أن المطلوب هو استرجاع الوثائق التى تحتوى على مصطلح الترويج Promotion ولكن استبعاد مصطلح الإعلان Advertising أى محاولة تجنب استرجاع وثائق غير متعلقة بالموضوع.

استبعاد أى وثيقة عن الإعلان وقبول أى وثيقة فيها الترويج فقط.

وهناك بعض المشكلات مع المنطق البوليني فى الاسترجاع ذلك لأن العديد من الأسئلة فى نظم استرجاع المعلومات، لا يتم التعبير عنها بوضوح أى أن السؤال يحتمل إجابات غير مرغوبة فمثلاً التعبير البوليني عن التعليم والفرنسى ومدارس يمكن أن

يسترجع تعليم اللغة الفرنسية فى المدارس الفرنسية، ويرى بعض الباحثين أن يكون العلاج فى التجمع الوثائقى document clustering (أى تجميع الوثائق المتشابهة) أو استخدام الوزن weight عن طريق التعرف على عدد مرات حدوث المصطلح فى قاعدة البيانات.

ثالثاً : أنواع الاسترجاع:

١- استرجاع الإشارات الببليوجرافية أو الاسترجاع المرجعى Reference Retrieval:

ويتمثل هذا النوع فى الفهرس البطاقى للمكتبة وغيره من الكشافات التى تدلنا على البيانات الببليوجرافية الكاملة عن الوثيقة سواء استخدم الكمبيوتر فى ذلك أو لم يستخدم.

٢- استرجاع نصوص الوثائق Document Retrieval: حيث يزود الباحث بالنصوص

الوثائقية الكاملة المطلوبة (أى تزويد الباحث بنسخة من الكتاب أو التقرير ..الخ) بدلاً من مجرد تزويد الباحث ببيان مرجعى عن الوثائق.

٣- استرجاع الحقائق أو المعلومات Data or Fact Retrieval: حيث يزود الباحث

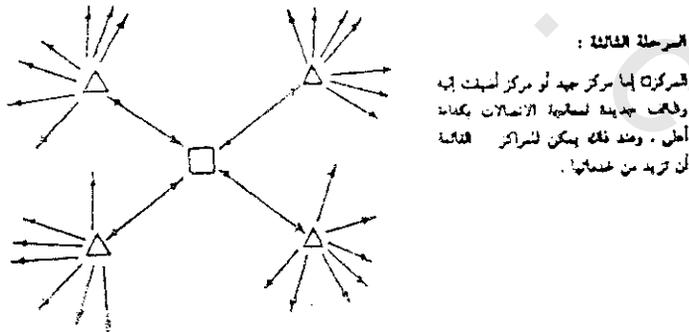
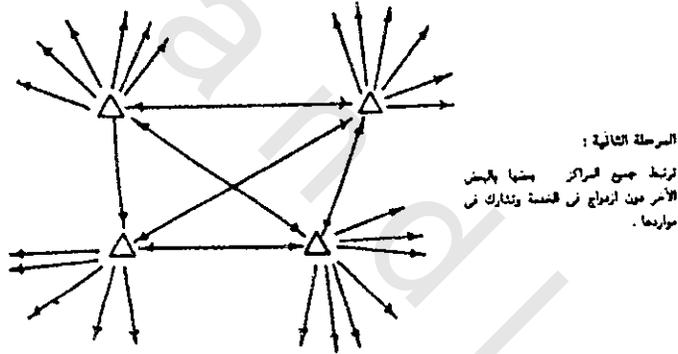
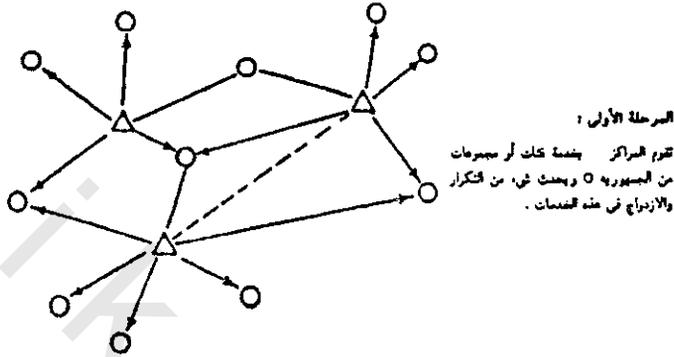
بالمعلومات نفسها المطلوبة كإجابة على أسئلة محددة (مثل سؤال عن الإنتاج الكلى لمحصول القطن فى السودان).

وتركز خدمات استرجاع المعلومات على المواد غير الكتنبية Material Non-Book أى أنها تهتم بالمقالات فى الدوريات والتقارير الفنية والمواصفات وبراءات الاختراع Patents وغيرها.

رابعاً : تقسيمات الشبكات وشبكة الانترنت:

كل واحدة من هذه التقسيمات لها ميزاتها وعيوبها، فالشبكة اللامركزية والتى قد تعكس أقل الإجراءات الإدارية هى نفسها أكثر تكلفة.. والتنظيمات المركزية مرغوبة ولكنها قد تكون معطلة إدارياً. ويمكن أن تعتبر الشبكات المركبة أو الهرمية ذات شكل توفيقى مقبول بين عدد الروابط Links المطلوبة وفرص التحويل، وإن كان المتبع فى معظم الأحوال هو تبنى توليفة الشبكة أى التى تأخذ من هذه الأشكال الأربعة، وذلك على ضوء التنظيمات والشبكات الموجودة فعلاً ودراسة الإمكانيات المادية والمشكلات القانونية وغيرها من العوامل.

ولقد وضعت مؤسسة نظم تدفق المعلومات في عام ١٩٧٨م التعريف التالي للشبكات، وهو الذي يتضمن ما يلي: يقصد بالشبكة اشتراك مؤسستين أو أكثر في تنظيم معين، يتيح لها المشاركة في المصادر والخدمات من خلال روابط الاتصال عن بعد، أى أن الشبكات مشروع تعاوني اقتصادي لمختلف الأطراف.



مراحل إنشاء شبكة المعلومات.

الشكل (١) بعض أشكال شبكات المعلومات والمكتبات

ولعل الشبكة العنكبوتية العالمية هى شبكة الشبكات والتي تضم عددا هائلاً من قواعد البيانات البليوجرافية والنصية والرقمية والتي يتم البحث عن طريقها وأهمها البحث بواسطة محركات البحث وأدلة شبكة الانترنت، وهذه تتميز بشمولية التغطية وحدثتها (حيث يظهر عادة آخر تاريخ تم فيه مراجعة هذا الموضوع على الويب) فضلاً عن إمكانية ربط المصطلحات البحثية فيها.

كما أن أدوات البحث تعمل على تجميع المعلومات وتكثيف الوثائق والبحث فى الكشافات وتعرض نتائج البحث، كما تقوم محركات البحث تقوم بمضاهاة المصطلحات طبقاً للسؤال المطلوب، وتتميز جوجل Google بواجهة بحث بسيطة حيث يتيح محرك البحث المتقدم البحث عن: مصطلحات أو صفحات معينة، كما يتيح استخدام محرك البحث بحيث يتم الاسترجاع باللغة المحددة وأحياناً ترجمة صفحات الويب ألياً إلى اللغة الإنجليزية من لغات متعددة.

وتعمل جوجل على توسيع نطاق خدماتها، حيث تخطط لما يسمى بالمكتبة العالمية حيث تداخلت مع اكبر المكتبات البحثية كجامعة أكسفورد بغرض تحويل الكتب المتوافرة لديها إلى الشكل الرقمي.

خامساً لغة التكشيف:

١/٦ مقدمة:

لا تعد لغة التكشيف مجرد قائمة من مصطلحات التكشيف المقبولة لدى المستخدمين، ذلك لأن لغة التكشيف تحتوى على آلية لبناء واستخدام هذه المصطلحات، أى إنها تشمل قواعد استخدام المصطلحات فى علاقة المصطلحات ببعضها، هذا وتخدم بعض المصطلحات فى إرشاد المستخدم للمصطلحات المضبوطة التى يمكن أن يستخدمها المكشف.

وهذه المصطلحات المؤشرة Pointer إلى جانب مصطلحات التكشيف المسموح بها تشكل ما يمكن تسميته بمصطلحات المداخل Entry Vocabulary أى أن مصطلحات المداخل الكلية تحاول التنبؤ بدقة - على قدر المستطاع - بجميع الطرق الممكنة التى توصل المستخدم بقاعدة المعرفة، ومصطلحات المداخل هذه تساعد كلاً من المستخدم

والمكشف على اختيار المصطلح الصحيح، وتقلل بالتالى من الجهد الفكرى اللازم لاتخاذ القرار .

ويمكن تقسيم لغات التشفيف إلى نظم مصطلحات معينة assigned n term systems ونظم مصطلحات مشتقة derived n term systems فى النظم الأولى يقوم المكشف بتعيين أو وضع مصطلحات أو واصفات على أساس التفسير الذاتى للمفاهيم الواردة فى الوثيقة وهو حين يفعل ذلك فهو يبذل بعض الجهد الفكرى، ذلك لأن المكشف يحدد موضوع الوثيقة كما يراه ثم يقرر ما هى المصطلحات المناسبة التى يتم ترشيحها. أما فى النظم المشتقة فجميع الواصفات تؤخذ من المادة نفسها.

وبالتالى فإن كشافات المؤلفين والعناوين وكشافات الاستشهادات Citations وكشافات اللغة الطبيعية تعد نظماً مشتقة، بينما تعد جميع لغات التشفيف ذات أدوات ضبط المصطلحات كقوائم رؤوس الموضوعات والمكانز وخطط التصنيف، هذه جميعاً تعد نظم مصطلحات معينة assigned، وهذه تتضمن جهداً فكرياً أكثر من غيرها، وبالتالى فهى مكلفة وتستغرق وقتاً أطول، أما النظم المشتقة فهذه يتم التفكير فيها على اعتبار إمكان القيام بها آلياً مثل كشافات كويك Kwic حيث يتم الاختيار بواسطة الحاسب الآلى، وقد تسمى التشفيف بالاقتراس Indexing by extraction.

وطبقاً لما يذهب إليه المدافعون عن النظم المعينة؛ فإن ضبط المصطلحات سيؤدى إلى انتظام التشفيف وكفاءته، ولكن الاختبارات العديدة التى أجريت على نظم ولغات التشفيف خلال السنوات الماضية لا تؤيد هذا الدفاع.

(Cleveland, D. 1990, 78)

سادساً : الحاجة إلى المصطلحات المحكومة :

هناك جوانب عديدة فى إعداد المصطلحات المحكومة، والإنتاج الفكرى حافل بالبحوث اللغوية فى هذا المجال، فقد أشار جوهانسن (Johansen, 1987) إلى أن إنشاء المصطلحات المحكومة يتطلب إنشاء علاقات بين الموضوعات ولا بد من التمييز هنا بين العلاقات المعتمدة على "اللغويات" والعلاقات المعتمدة على تميز "اللغة" وقدم منهجاً للتعرف على بناء الموضوع عن طريق اختبار عناصره المكونة وعلاقات أجزائه

بعضها ببعض فضلاً عن تناول الموضوع من ناحية العلاقة التركيبية Syntactic فى لغات التكشيف.

ويعد استرجاع المعلومات فى الإنسانيات بصفة عامة أكثر صعوبة منه فى العلوم، وذلك لأن المصطلحات المستخدمة فى الإنسانيات تعد أقل دقة؛ ويتناول الباحث ويبرلى (Wiberley, 1983) ذلك عن طريق دراسة المداخل المستخدمة فى الموسوعات والقواميس المشهورة فى مجال الإنسانيات، ويخلص من دراسته المشهورة فى مجال الإنسانيات، إلى أنه على الرغم من أن هناك بعض المصطلحات غير الدقيقة والمحددة imprecise إلا أن معظمها دقيق ويحتوى على أسماء أفراد أو أعمال إبداعية... ولعل هذه النتائج تشير إلى أن الإتاحة الموضوعية فى الإنسانيات يمكن ألا تكون مشكلة جوهريّة؛ وهناك الأهمية الكبرى لمعرفة اللغات نفسها، ذلك لأن هذه المعرفة تؤدى دوراً مهماً فى اختيار مصطلحات التكشيف والتحكم فى المكائز والإحالات التبادلية Cross-References فضلاً عن نقاط الإتاحة، ويشير الباحث إلى أن تطويع النص فى قواعد المعلومات العالمية يتطلب نوعاً من المعرفة بثلاث لغات Trilingualism وهذه تشمل معرفة بلغة المستفيد الوطنية، ومعرفة باللغة الأجنبية الصادر بها البحوث، ومعرفة الجوانب الثلاثة سيؤدى إلى تقليل أداء النظام، ويخلص الباحث من ذلك إلى أهمية دراسة اللغات الأجنبية إذا أريد خدمة المستفيدين بطريقة أفضل.

وهناك مفهومان متعلقان بضبط المصطلحات وهما مفهوم السند الأدبى Literary Warrant ومفهوم سند المستفيد User Warrent فمصطلح التكشيف فى كتاب أو فى قائمة ضبط المصطلحات يدل على أن له سنداً أدبياً إذا كان هناك إنتاج فكرى كاف عن الموضوع وممثلاً بهذا المصطلح... أما سند المستفيد فهو مفهوم مشابه، فيقال إن مصطلحاً معيناً يعد سنداً Warranted إذا كان هذا المصطلح مستخدماً بطريقة منتظمة بواسطة المستفيدين للكشاف فى بحثهم عن المعلومات... ومن الواضح أن سند المستفيد يعد مفتاحياً بالنسبة لتصميم أداة ضبط المصطلح. كما يتطلب أن يكون المصمم مدركاً لما يحتاجه المستفيدون وعلى كل حال فالمصطلحات المضبوطة لها صفات مميزة كما يلي:

- ١- تمثل هذه المصطلحات البناء المفهومى العام للمجال الموضوعى، كما تقدم لنا مرشداً للمستفيد من الكشاف.
- ٢- تشتق المصطلحات من تلك المستخدمة أى إنها تعكس مصطلحات الإنتاج الفكرى وتعكس الاستخدام الفنى للمستفيد... وعلى الرغم من أن هذه المصطلحات لا تستطيع أن تكون مخصصة Specific كما فى كشف اللغة الطبيعية؛ إلا أنها يجب أن تتيح إمكان الدقة للاسترجاع الجيد.
- ٣- تستخدم هذه المصطلحات المضبوطة عدداً مناسباً من الحمل سابقة التجهيز حتى تقلل من فرص الحصول على نتائج خاطئة False drops على قدر المستطاع.
- ٤- تزودنا بمصطلحات معيارية عن طريق التحكم فى المترادفات وبالتالي تزيد من إمكان الانتظام، أى إنها تزودنا بآلية تسمح لنا باستخدام مصطلح واحد فقط من قائمة المصطلحات المشابهة لكشف مفهوم معين.
- ٥- يمكن أن تعرف لنا المصطلحات الغامضة، كما يمكن أن تميز بين مصطلحات الجنس Hemographs (تماثل فى اللفظ واختلاف فى المعنى).
- ٦- عن طريق الإحالات يمكن أن تظهر لنا العلاقات الأفقية والرأسية بين المصطلحات.

وعلى الرغم من الإمكانيات المحتملة الكبيرة لاستخدام اللغة الطبيعية فى الحاسب الآلى بواسطة المستفيد الذى يفيد من المصطلحات المضبوطة أو المحكومة أو أنه يفيد من نظام لغة طبيعية محدود. ذلك لأن المصطلحات المحكومة ستعالج المترادفات وترتبط بين المصطلحات المتقاربة دلاليًا، وبالتالي تقلل عملية تخمين الكلمة الصالحة لدى المستفيد.

سابعاً: توليد المصطلحات وزيادة كفاءة التعبير والنمو:

هناك طريقتان لتكوين وبناء المصطلحات، الأولى ما يمكن تسميته بالمصطلحات التطويرية revolutionary vocabulary وهذه تتكون كنتيجة لتوقف وضع المصطلحات الحرة عند نقطة معينة ثم محاولة وضع تنظيم معين لهذه المصطلحات... أما الطريق الثانى فهو الذى يتم فيه توليد المصطلحات كنتيجة لدراسة متخصصة واتفاق بين

الخبراء الذين يحددون ماذا ينبغي أن يدرج من المصطلحات في مجال معين من المعرفة ... وهذه الأخيرة تسمى المصطلحات الحصرية Inumerated vocabulary.

ونحن نلاحظ أن المصطلحات التطويرية تتكون من مواد خام يقدمها المكشفون، وبعد تكشيف عدد كافٍ من الوثائق يتم وضع المصطلحات في ترتيب هجائي بغرض التحرير واتخاذ إجراءات القبول للمصطلح بعد تعريفه داخل مجال معين [مثال كلمة قاعدة (عسكرية) أو قاعدة (في الرياضيات) أو قاعدة (في الكيمياء) ... الخ]، هذا بالإضافة إلى تحديد المرادفات ومعناها ثم وضع المرادف المفصل Preferred Term ويتم ضبط المصطلحات المترادفة بواسطة قاموس.

وليس هناك لغة قادرة على إعطاء مستوى الاتساع أو التحديد في نقطة معينة، بمصطلح معين، وبالتالي يلجأ الباحث أو المكشف إلى توليفة من الواصفات لزيادة كفاءة التعبير عن المطلوب، وهذه تتم عادة بالتوليفات البولينية Boolean Combinations المستخدمة في نظام الترابط اللاحق.

والآن لابد من الإشارة إلى إحدى خصائص نمو المصطلحات الكشفية - سواء الحرة أو المضبوطة، أي إنها دائمة التغيير وذات صفات ديناميكية ... وبالتالي فإن الإنتاج الفكرى فى موضوع معين يتغير، وبالتالي فإن لغة التكتشف التى تمثل محتويات الموضوع ستكون أيضاً ديناميكية متغيرة؛ والمصطلحات لا تتغير وحدها ولكن حجم الوثائق ينمو والموضوعات القديمة تنتقسم أو تنضم لتشكيل مجالات جديدة [البيولوجيا والكيمياء تصبح كيمياء حيوية]، أى إن المصطلحات القديمة لم تعد صالحة أو كافية ومع ذلك فبقاؤها ضرورى لمسايرة التفكير والوضع القديم، أى إن لغة التكتشف تنمو بالضرورة مع ما يصحب ذلك من تعقيدات بنائية... (Cleveland, D. 1990, 86-88).

ثامناً : بعض جوانب المقارنة بين اللغة الطبيعية والمصطلحات الحكومية :

يمكن تتبع الاهتمام ببحث النص بالحاسب الآلى إلى أوائل الخمسينات، ولكن الاهتمام بهذا الموضوع استمر فى الزيادة مع توافر النص (المستخلصات أو المواد الكاملة) الذى يمكن إتاحتها على الخط المباشر.

ولابد هنا من التنويه إلى أن النص الممكن إتاحتة على الخط يتكون من مواد قصيرة كمقالات الدوريات والمستخلصات وبراءات الاختراع ... ونظراً لانخفاض التكاليف السريع (بسبب توافر الأقراص المكتزة مثلاً) فقد أصبح من الممكن اختزان وتطوير النص لمواد أكثر شمولاً ... ولكن إلى أى مدى سنكون قادرين على بحث النص الكامل ... للكتاب مثلاً ؟

وإذا قمنا بمقارنة نتائج البحوث فى القواعد البليوجرافية المكشفة بواسطة الإنسان، بنتائج البحوث فى القواعد البليوجرافية المكشفة بواسطة الإنسان، بنتائج البحوث فى قواعد البيانات كاملة النص؛ أن الأخيرة فى قواعد البيانات كاملة النص؛ أن الأخيرة تقدم لنا استدعاءً أكبر greater recall ولكن دقة أقل reduces precision ويشير العديد من الباحثين إلى أن ذلك يعود إلى:

أ- طول التسجيلة أو المستخلص.

ب- تقليص عدد الواصفات المخصصة.

وقد يؤدى بحث النص الكامل فى أحيان كثيرة إلى إظهار مزيد من التفاصيل أما البحث بواسطة المصطلحات المحكومة فيميل إلى إظهار مفاهيم واسعة ... أى إن البحث الشامل يتطلب استخدام أكثر من طريقة واحدة ...

وينبه بعض الباحثين إلى محاذير عند وجود قواعد بيانات النص الكامل بطريقة عادية، وميسرة، ذلك لأن مجرد وجود النص الكامل قد يوهم بأنه قد يكون البديل الذى يوفر على المكشفين مهمتهم فى التكشيف الموضوعى. وقد يرى المزج بين النص الكامل وأساليب البحث الموضوعى بديلاً مفضلاً للاسترجاع الأفضل.

وهناك بعض العوامل الرئيسية التى تؤثر على أداء نظم استرجاع المعلومات (كما هو الحال فى الشكل التالى فىمكن أن نرى على الجانب الأيمن ثلاث تمثيلات للوثيقة من النص الحر (وهى العنوان واثنان من المستخلصات مختلفى الطول). أما على الجانب الأيسر - فهناك مجموعتان من المصطلحات الكشفية (التي تغطى الجانب الموضوعى بطريقة انتقائية وشاملة)، والمصطلحات تعتمد على مكنز الأمم المتحدة UNBIS، ولعل أحد هذه العوامل الرئيسية التى تؤثر على أداء نظم استرجاع المعلومات

هو عدد نقاط الوصول المتوافرة، ومن الواضح أن المستخلص الموسع يقدم لنا نقاط الوصول أكثر من المستخلص المختصر، وهذا بدوره يقدم نقاط وصول أكثر من العنوان، وقياساً على ذلك فإن التشفيف الموسع يقدم لنا نقاط وصول تفوق التشفيف الانتقائي فبحث النص على العنوان وحده - ربما يسترجع لنا المادة التي يعد موضوعها المحورى مركزاً فى العنوان وهكذا - عند زيادة نقاط البحث، حيث يمكن التعرف على جوانب إضافية.

- وهناك جانب آخر فى عملية الاسترجاع هو طول التسجيلة نفسها، وليس نوع المصطلحات المستخدمة.

- ويمكن تحليل هذا الوضع ورؤيته بطريقة أكثر عمقاً عندما نختبر قاعدة معلومات تحتوى على آلاف المواد المكشفة بطريقة شاملة (كما هو الحال فى المثال الموجود فى الشكل التالى فمثل هذه القاعدة توفر لنا استدعاء Recall أعلى من قاعدة بيانات توفر لنا الوصول من خلال العناوين فقط، وكذلك فإن قاعدة المعلومات التى تحتوى على مستخلصات موسعة ستوفر لنا درجة استدعاء أعلى من ذلك الذى يعتمد على التشفيف الانتقائى أو حتى التشفيف الشامل، وهذه المقارنة لا تتصل بمقارنة اللغة الطبيعية بلغة المصطلحات المحكومة فى استرجاع المعلومات، ولكنها تتعلق فقط بطول التسجيلة القابلة للبحث.

وخلاصة هذا كله أن المستخلصات ستزودنا بنقاط وصول أكثر من مجموعة الواصفات التى يضعها مكشف بعينه، كما أن النص الكامل للوثيقة سيؤدى بنا إلى نقاط وصول أكثر أيضاً، وبالتالي نحن نتوقع أن قاعدة معلومات النص الحر ستقدم لنا استدعاء أكبر من ذلك الذى يعتمد على التشفيف الانسانى.

- وهناك عامل آخر يؤثر على أداء نظام الاسترجاع وهو التخصيص Specificity، ذلك بالنسبة لوصف الجوانب الموضوعية للوثيقة، ويلاحظ هنا أن مصطلحات مكنز الأمم المتحدة UNBIS هى مصطلحات مخصصة للغاية لوصف معظم جوانب الموضوعات التى تظهر فى الشكل رقم (1)، ومع ذلك فإن النص الحر يقدم لنا درجة تخصيص أكبر.

الشكل رقم (٣)

تأثير طول التسجيلة على الاسترجاعية

العنوان:	تكشيف (انتقائي)
مسح الرأي العام الوطني لاتجاهات الولايات المتحدة عن الشرق الأوسط.	- الرأي العام.
مستخلص (مختصر):	- المسوحات الهاتفية
يقدم مسح تم عن طريق الهاتف عام ١٩٨٥م وجهات النظر بالنسبة لبعض القضايا مثل: المعونة الأمريكية لإسرائيل ومصر؛ وهل تتخذ أمريكا موقفاً مالياً لإسرائيل أم للدول العربية أو موقفاً حيادياً؟ وهل يجب على منظمة التحرير الفلسطينية المشاركة في مؤتمر للسلام، وهل تعد الدولة الفلسطينية المستقلة مطلب سابق للسلام؟	- الولايات المتحدة الأمريكية
مستخلص (موسع):	- الاتجاهات
تم القيام بمقابلات هاتفية عام ١٩٨٥م مع عدد [٦٥٥] أمريكياً وكانت العينة احتمالية، وقد تم الحصول على إجابات للأسئلة التالية: هل إنشاء دولة فلسطين أساس للسلام؟، هل يجب تقليل المعونة الأمريكية لكل من إسرائيل ومصر؟، هل يجب مشاركة أمريكا في مؤتمر للسلام بحيث يضم المؤتمر منظمة التحرير الفلسطينية؟ هل يجب على أمريكا عدم الوقوف إلى جانب إسرائيل أو إلى جانب الدول العربية بل عليها أن تحتفظ بعلاقات صداقة مع الاثنين؟... ولقد تم التعبير عن الآراء بواسطة عدد من قادة الشرق الأوسط الرئيسيين (حسين، عرفات، بيريز، مبارك، فهد، الأسد) خصوصاً بالنسبة لجهودهم من أجل السلام، فضلاً عن محاولة التعرف من المستجيبين إذا كان لديهم معلومات كافية عن الجماعات الوطنية المختلفة بالمنطقة.	- الشرق الأوسط
	- إسرائيل
	- مصر
	- الدول العربية
	- فلسطين
	- منظمة التحرير
	- مؤتمرات السلام
	- السلام
	- الدولة الفلسطينية
	- المعونات الخارجية
	- الزعماء السياسيون

حيث يوفر لنا استرجاع أسماء قادة الشرق الأوسط مثلاً، بينما التكشيف يسمح لنا فقط بالبحث على مستوى القادة السياسيين، ومعنى هذا كله أنه كلما زادت نقاط الوصول Access Points التي تساعدنا على الاسترجاع كان الاستدعاء Recall أعلى، وفي الوقت نفسه نستغل الدقة Precision، والسبب في ذلك ببساطة أنه كلما زادت نقاط الوصول، فإن بعض هذه النقاط سيتصل بجوانب ثانوية للوثيقة، ونتيجة لذلك فإن

طالب السؤال الذى يتلقى الاستجابة عن مادة فى الشكل رقم (١)، وبالنسبة لبحث عن القائد الفلسطينى عرفات، سيصل إلى نتيجة مفادها أن هذه المادة لا فائدة منها لأنها تتعلق بالقائد السياسى عرفات بطريقة مختصرة غير وافية.

وهناك عيب آخر يأتى من زيادة نقاط الوصول، وهى إمكان وجود بعض العلاقات ملتبسة Spurious وقد تكون هذه العلاقات فى شكلين:

أ- الارتباطات الزائفة False association.

ب- علاقات مصطلحات غير صحيحة.

ويمكن أن يعكس لنا الشكل السابق إمكانات عديدة فى هذا الشأن، فالمستخلص الموسع قد يودى إلى استرجاع مادة من خلال الذى تم فى مقابلة هاتفية مع قادة الشرق الأوسط (أو أى من هؤلاء القادة المذكورين. كما أن الكشف الموسع قد يودى إلى استرجاعها فى بحث عن القادة السياسيين بالولايات المتحدة الأمريكية. فمثل هذه العلاقات والارتباطات هى بالضرورة ارتباطات أدت إلى استرجاع لا علاقة له بالوثيقة (المقابلة الهاتفية لم تكن تتصل مباشرة بالقادة السياسيين، كما أن الولايات المتحدة الأمريكية ليست متصلة مباشرة بالقادة السياسيين).

وقد تظهر هذه العلاقات غير الدالة فى الكشف الانتقائى أو حتى عند استخدام العنوان، وفى هذه الحالة فإن المصطلحات التالية (الشرق الأوسط - الاتجاهات - الولايات المتحدة الأمريكية) مصطلحات مترابطة فيما بينها، ولكن العلاقة غامضة.

والكشف الإنسانى هو عملية فكرية ذاتية، أى إن التميز الشخصى قد يدخل فى العملية، وبالتالي سيضيف المكشف موضوعات لا داعى لها أحياناً، كما قد يستخدم مصطلحاً لا يمثل الموضوع تمثيلاً سليماً، كما أن المكشف قد لا يظهر العلاقات ذات الأهمية لمستفيدين بعينهم، ولعل النص الكامل قد يجنبنا هذا النوع من المشكلات، ولقد اهتدى الباحث (Horty 1982) إلى هذه الحقيقة منذ أكثر من عشرين عاماً حين قال: عندما يستخدم النص الكامل للوثائق كأساس لنظام الاسترجاع، فإن سؤال الباحث لن يكون فى إجابته محدوداً بالطريقة التى يتم بها كشف الوثائق، فقد يتطرق الأمر إلى بعض الموضوعات المستقبلية، وقد يكون ذلك على قدر كبير من الأهمية، وهذا

لا يتوقعه المكشف فى معظم الأحيان. وبالتالي فالاستغناء نهائياً عن الكشف، والبحث المباشر فى النص الأصلي قد يعطى نتائج أفضل. وهناك جانب آخر لا بد من الإشارة إليه وهو أن هذه المشكلة الخاصة بالمصطلحات غير المناسبة فى وثيقة واحدة تؤدي إلى علاقات زائفة فى الاسترجاع، هذه المشكلة تتضاعف وتتزايد مع تعدد الوثائق ومع تجميعها لتشكيل قاعدة بيانات ضخمة.

والربط الدلالي بين المصطلحات المتعلقة ببعضها semantically سيساعد الباحث على تحديد جميع المصطلحات التي يحتاجها للوصول إلى بحث كامل.

- وهناك عامل إضافة مصطلحات جديدة تدخل فى العناوين أو المستخلصات، هذه المصطلحات لن تظهر فى المكانز وقوائم المصطلحات المحكومة إلا بعد مدة طويلة، وبالتالي فإن اللغة الطبيعية تعد ذات أهمية بالنسبة للموضوعات الجديدة وبالتالي ستحسن كل من الاستدعاء والدقة، وأخيراً فإن المصطلحات المحكومة مفضلة لدى اختصاصى المعلومات الذى سيصبح بعد مدة قادراً على السيطرة على لغة البحث، بينما ستكون اللغة الطبيعية عادة هى المفضلة بالنسبة للمتخصص الموضوعى.

لعل المناقشة السابقة قد حددت بعض الخصائص المميزة للنص الحر وللتكشيف الإنسانى باستخدام المصطلحات المحكومة، وعلاقة ذلك بكل من الاستدعاء والدقة، وتتضح هذه العلاقات ويتم تلخيصها فى الشكل رقم (٢).

وواضح من هذا التحليل السابق أن الأمر ليس يسيراً، فهناك بعض العوامل التى تؤيد المصطلحات المحكومة، بينما يؤيد بعضها الآخر النص الحر، فتخصيص كلمات النص يميل إلى تحسين عامل الدقة، ولكن ذلك يجعل الوصول إلى استدعاء عالٍ أمراً عسيراً، على الأقل بالنسبة لبحوث المفاهيم العريضة، بينما يؤدي طول النص إلى تحسين الاستدعاء، ولكنه يقلل الدقة فى الوقت نفسه. ومع ذلك فلا نستطيع تفضيل هذا على ذلك، أو على العكس، لأن التفضيل هذا يعتمد على نوع البحث. فالبحث الذى يتضمن المفاهيم العريضة يفضل معه المصطلحات المحكومة، أما البحث الذى يتطلب التخصيص العالى (خصوصاً ذلك الذى يتضمن أسماء أفراد أو مؤسسات ... الخ) سيحبذ النص الحر، أما البحث الشامل الحقيقى عن موضوع معين (مثلاً جميع

المصادر الممكنة عن دواء معين)، هذا البحث يحبذ استخدام النص الكامل، وأخيراً فالبحث الانتقائي (الذى يتطلب المواد الأكثر أهمية)، فهو يحبذ التكشيف بالمصطلحات المحكومة.

ويمكن بصفة عامة أن يقال إن البحث باللغة الطبيعية يمكن أن يؤدي إلى نتائج أفضل بالنسبة للمفاهيم المحددة، وليس بالنسبة للمفاهيم العام (fugmannl 1985) أما الباحث ديبوا (Dubois 1987) فيذهب إلى أن أحد مزايا النص الحر، وهو إمكان استيعاب المصطلحات الجديدة. أما الباحث بيريز (Perez 1982) فيقرر أن المصطلحات المحكومة قد تؤدي إلى عدم الوصول إلى الدقة Precision، بينما لا يفقدنا النص الحر الموضوعات التخصصية.

كما ينبغي أن يؤخذ في الحسبان جانب التكاليف عند مقارنة اللغة الطبيعية بالمصطلحات المحكومة فتكاليف التجهيز الفكرى الإنسانى تزيد بصفة مستمرة وبمعدلات عالية، وذلك عند مقارنتها بتكاليف التجهيز الآلى، كما أن التكشيف بالمصطلحات المحكومة يتطلب عمالة كثيفة، فضلاً عن أن إنشاء ومتابعة المصطلحات المحكومة يعد أمراً مكلفاً، ومن جهة أخرى فنصوص المقالات والبحوث تصبح على مر الأيام متاحة بتكاليف زهيدة فى الشكل المقروء آلياً، وذلك كنتاج إضافى لمناشط النشر والبث.

ومن هنا أصبح من الأمور الطبيعية أن يقيم مدير خدمات المعلومات الموقف بعناية لتحديد مدى ميزات التكشيف بالمصطلحات المحكومة، وهل تعد التكاليف الإضافية المذكورة سابقاً مبرراً كافياً لاستمرار استخدامها؟

فالاستغناء عن التكشيف الذى يقوم به الإنسان باستخدام المصطلحات المحكومة، سيؤدي غالباً إلى تخفيض تكاليف المدخلات، ومع ذلك فإن ذلك يتم على حساب تكاليف كبيرة للمخرجات، على اعتبار أنه سيكون هناك عبء فكرى متزايد على المستفيدين من قواعد المعلومات، ولعل هناك بعض العوامل الأخرى التى ستؤثر على قرار المخرجات السابقة، ومن بينها حجم الوثائق والبحوث وتكاليف التكشيف والبحث ودرجة الأهمية التى يمكن أن نردها لنتائج البحث.

الشكل رقم (٤)

المزايا والعيوب للنص الحر مقارنة بالمصطلحات المحكومة

أثر هذا النوع من التمثيل	* العوامل التى تساعد على الاستدعاء
معظم تمثيلات النصوص الحرة ستكون عادة أطول من المجموعة المعينة من مصطلحات التكثيف. وهذا من شأنه أن يؤدي إلى الاستدعاء لكنه يقلل الدقة (وهناك زيادة فى حالات العلاقات الزائفة).	أ- طول التسجيلة (عدد نقاط الوصول)
هذه الخاصية ستكون عادة أكبر فى النص الحر، وبالتالي ستحسن من إمكان وجود مادة معينة، ومع ذلك فإن الطرق المختلفة الكثيرة التى يمكن التعبير بها عن موضوع معين من قاعدة بيانات نصية ضهمة يجعل من العسير الوصول إلى استدعاء عال.	ب- التكرار غير المرغوب فيه (الحشو) Redundancy
وهذه تظهر عادة فى تمثيل المصطلحات المحكومة. وهذا واضح فى المصطلحات المبنية بناءً سليماً (BT, NT, RT)	ج- وجود مفاهيم ومصطلحات عريضة. د- ربط المصطلحات دلاليًا Semantically
النص الحر يكون عادة أكثر تخصيصاً، أى إنه يؤدي إلى الدقة المطلوبة، ومع ذلك فتتويع الطرق التى يتم التعبير بها عن المفاهيم يجعل من العسير جداً الوصول إلى استدعاء عال فى بحوث المفاهيم العريضة، وفى مثل هذه البحوث فالمصطلحات العريضة المحكومة نسبياً ستكون المفضلة.	* العوامل التى تساعد على الدقة: - التخصيصية Specificity
ستكون تمثيلات النص الحر دائماً أكثر مواكبة، ذلك لأن العثور على موضوع جديد فى نظام المصطلحات المحكومة يتطلب من الباحث تجريب عدد من المصطلحات الأخرى (وبالتالى تقليل الدقة) ومع ذلك فسوف لا يعثر على كل شئ عن هذا الموضوع (أى تقليل الاستدعاء). اختصاصى المعلومات يكون أكثر ألفة بالمصطلحات المحكومة ويستخدمها بطريقة أكثر فاعلية من غيره، أما المستفيد النهائى، فسيكون أكثر ألفة باللغة الطبيعية الموجودة فى الوثائق فى مجاله الموضوعى المحدد.	* عوامل تؤثر على كل من الاستدعاء والدقة: أ- المواكبة Currency ب- الألفة Familiarity

وخاصة ذلك كله أن اللغة الطبيعية والمصطلحات المحكومة لها ميزاتهما ولهما عيوبها، فتسجيلات النصوص الحرة تميل إلى أن تكون أطول من غيرها، وبالتالي تزودنا بنقاط وصول أكثر، وستحتوي بعض المصطلحات الأكبر تخصيصاً أو الأكثر حداثة من تلك المصطلحات المحكومة، أما المصطلحات المحكومة من جانب آخر، فهي تفرض عملية الانتظام أى تمثيل المحتوى الموضوعي.

تاسعا: تحسين الاسترجاع عن طريق تحسين عمليات ولغات الأسئلة وتحسين عمليات ولغات النص:

تغطي هذه الدراسة وسائل تحسين الاسترجاع عن طريق لغات وعمليات السؤال وعن طريق لغات وعمليات النص.

١/١١ تحسين لغات الأسئلة:

تغطي هذه الدراسة الأنواع المختلفة من الأسئلة التي يجب عليها النص في نظم الاسترجاع وتعتمد هذه جزئياً على نموذج الاسترجاع الذي يتبناه على النظام المعتمد على ترتيب الكلمات المفتاحية (مثل محركات بحث الوب) أو على نموذج الهيبرتكتست. ويضم هذا لتحسين التعرف على كيفية صياغة الأسئلة، ذلك لأن نوع السؤال الذي يمكن أن يقوم بصياغته الباحث يعتمد إلى حد كبير على نموذج استرجاع المعلومات الذي وراءه.

ولابد هنا من التمييز بين استرجاع المعلومات واسترجاع البيانات لأننا سنستخدم هذه الثنائية في تصنيف لغات الأسئلة المختلفة، وقد اخترنا أن نميز بين اللغات التي تسمح بترتيب الإجابة وهي لغات الأسئلة المختلفة، وقد اخترنا أن نميز بين اللغات التي تسمح بترتيب الإجابة وهي لغات لاسترجاع المعلومات (النماذج الأساسية لاسترجاع المعلومات) حيث يعتبر الاسترجاع المعتمد على الكلمات المفتاحية هو النوع الرئيسي لمهمة السؤال.

أما بالنسبة للغات السؤال التي لا تهدف إلى استرجاع المعلومات فإن مفهوم الترتيب Ranking لا يمكن تحديده بسهولة، وبالتالي فنحن نعتبر أن بعض أسئلة اللغات لا يقصد منها أنها نهائية بالنسبة للمستفيد ويمكن أن ترى كلغات ذات مستوى عالٍ من

البرمجة ويتم السؤال هنا لقاعدة البيانات على الخط المباشر أو فى أرشيف الاقراص المكتتزة CD-ROM وفى هذه الحالة نحن نتحدث عن بروتوكول وليس على لغات الأسئلة. واعتماداً على خبرة المستخدم تستخدم لغة أخرى للسؤال، عمل الاسترجاع والترتيب فى هذه الحالة ربما لا يتم الاحتياج إليه.

والقضية الهامة أن معظم لغات السؤال تحاول استخدام المحتوى (أى المعانى والدلالات Semantics وتركيب Structure النص (أى النحو بالنسبة للنص) للعثور على الوثائق المتعلقة Relevant من أجل ذلك فإن النظام قد يفشل فى إيجاد الإجابات المتعلقة وبالتالي لا بد من استخدام عدد من الأساليب التى تهدف إلى تسريع فائدة الأسئلة الموجودة، ومن أمثلتها توسيع الكلمات إلى مجموعة من المترادفات أو استخدام المكنز Thesaurus واستخدام جذور الكلمات Stemming بالإضافة لمختلف مشتقات نفس الكلمة. فضلاً عن استبعاد الكلمات التى لا تحمل معنى (مثل كلمة The) والتى تسمى Stopwords أى الكلمات المحذوفة.

وقد نظم هذا الفصل كما يلى فى الجزء الأول نذكر الأسئلة التى يمكن صياغتها بالكلمات المفتاحية التى تعتمد على لغات السؤال وهذه تحتوى على تطويع المشغلات البولينية.

وفى الجزء الثانى نقوم بتغطية نموذج المضاهاة التى تحتوى على أسئلة أكثر تعقيداً والتى تهدف إلى التكامل مع البحث بالكلمات المفتاحية ببيانات أكثر قوة وإمكانية استرجاعية. وثالثاً نحن نغطى الأسئلة الخاصة بتركيب النص والذى يكون أكثر اعتماداً على نموذج النص Text model وأخيراً نحن ننتهى ببعض البروتوكولات المستخدمة على الانترنت (Wide Area I sevice WAIS) بواسطة الناشرين للأقراص المكتتزة.

لقد تمت فى هذا الفصل التعرف على الأوجه الرئيسية للغات الأسئلة والتى يمكن أن نسترجع المعلومات من قواعد البيانات النصية، وغطت المناقشة معظم الأدوات التقليدية الامكانيات الجديدة من الكلمات البحثية إلى النماذج أو المثل الممتدة ومن نموذج البوليان إلى تركيبات الأسئلة.

خدمات الاستخلاص والضبط الببليوجرافي للكشافات والمستخلصات

لقد بدأت خدمات الاستخلاص لمعاونة العلماء الباحثين في مسابرة النمو التضاعفي في الإنتاج الفكرى العلمى وتعتبر مستخلصات العلوم Sciene Abstracts التى بدأت عام ١٨٩٨ من أوائل خدمات الاستخلاص وانقسمت مستخلصات العلوم منذ عام ١٩٠٣ إلى مستخلصات الفيزياء physics Abstracts والمستخلصات الهندسية الالكترونية والكهربائية والمستخلصات الضابطة والمحسبة Computer & Control Abstracts أما المارد فى هذه الفئات العلمية فكان المستخلصات الكيميائية Chemical Abstracts التى بدأت عام ١٩٠٧ بعدد مستخلصات بلغ أكثر من ١٧٠,٠٠٠ مقال مستخلص من عدد (١٢٠٠٠) دورية... وعلى كل حال فقد وصل عدد خدمات التكشيف والاستخلاص حالياً إلى حوالى (١٥٠٠) خدمة.

والعديد من هذه الخدمات متاحة على الخط المباشر وبعض هذه الخدمات قد أوقف إصدار المستخلصات المطبوعة واكتفى بتلك المذاحة على الخط المباشر.

ذلك لأن المزايا المتوفرة فى الخدمات على الخط المباشر هى خدمات تفوق تلك المطبوعة سواء بالنسبة للخدمة الكشفية أو خدمات الاستخلاص، وهناك ثلاثة أنواع من المستخلصات هى المستخلصات الشارحة Indicative Abstracts والمستخلصات الإعلامية Informative Abstract أما الثالثة فهى التى تجمع بين الاثنين والمستخلصات الشارحة هى التى تشبه قائمة المحتويات ولكن المستخلصات الإعلامية تهدف إلى تقديم تفاصيل أكثر وبعض هذه المستخلصات يمكن أن يكون بديلاً للمقالات الأصلية إذا وفرت تفصيلاً كافياً، كما أن بعضها يمكن أن يتخطى حاجز اللغة أى أنها مثلاً تصدر باللغة الانجليزية بما فى ذلك ترجمة هذه المستخلصات (باللغات الأخرى) للغة الانجليزية.

أما بالنسبة للضبط الببليوجرافى لهذه الكشافات والمستخلصات فهناك الدليل العالمى لأولرخ Ulrich's International periodicals Directory الذى ينشره باوكر Bowker وكذلك هناك مرجعين هاملين أحدهما يصدر فى بريطانيا وهو Walford Guide to Reference Material والذى تنشره جمعية المكتبات البريطانية والآخر Sheehy Guide to Reference Books وتنشره الجمعية الأمريكية للمكتبات (ALA).

وأخيراً بالنسبة لخدمات التكشيف ولاستخلاص فى مجال علم المكتبات والمعلومات فهناك ثلاثة مراجع أساسية هى يغطى:

أ- الإنتاج الفكرى للمكتبات (وعلم المعلومات) Library Literature للناشر ولسن Wilson وهو كشاف بالمؤلفين والموضوعات وهو يغطى مختلف الإصدارات من الكتب والأطروحات وهو متاح حالياً على الخط المباشر (www.hwwilson.com).

ب- مستخلصات علم المكتبات والمعلومات Library & Information Science Abstracts (LISA-1969) والتي يصدرها Bowker وتضم مقالات من حوالى (٥٠٠) دورية وبعض المؤتمرات من حوالى (٦٠) دولة وتقوم مستخلصات باللغة الانجليزية من حوالى (٣٤) لغة أجنبية.

ج- مستخلصات علم المعلومات (1966) Information Science Abstracts وتقوم مستخلصاً من حوالى (٧٠٠) دورية ومعظمها من الإنتاج الفكرى التكنولوجى حيث تضم النظم المحسبة والبرامج أى أنها ملائمة للمتخصصين فى نظم المعلومات وهناك نسخة متاحة على الديالوج وسيلفر بلانتر (مقتبسات من: الموسوعة الدولية لعلم المعلومات والمكتبات IEILS, 1-3, Chandler, Helen (2003)) ويتوفر فى الوقت الحاضر بوابة عربية للمكتبات والمعلومات وهى: www.cybrariansinfo

مراجع الفصل السابع

المراجع العربية :

- ١- أحمد أنور بدر (أكتوبر ١٩٩٧) اللغة الطبيعية والمصطلحات المحكومة فى إسترجاع المعلومات مجلة مكتبة الملك فهد الوطنية، مج ٣، ع ١ ص ١٠١-١٤٤٤.
- ٢- أحمد أنور بدر (أبريل ٢٠٠٩) الإنتاج الفكرى لمجال إسترجاع المعلومات وواقع تدريسه فى بعض أقسام المعلومات والمكتبات المصرية والسعودية: دراسة توثيقية ميدانية. الاتجاهات الحديثة فى المكتبات والمعلومات، مج ١٦، ع ٣٢، ص ١١-٥٤.

المراجع الأجنبية :

- 3- Brown, K. et al (1990) A Statistical * approach to machine translation. *Computational Linguistics*, v.16 (2), p. 77-85.
- 4- Chen, H; Yim, T.; Fye, D. (1995) Automatic Thesaurus Generation for an electronic Community System. *Journal of the American Society for Information Science*, v. 46 (3), p. 175 - 193.
- 5- Cleveland, D.B. and Cleveland, A.D. (1990) *Introduction to Indexing and Abstracting*. 2nd ed. Englewood Colorado: libraries Unlimited. Inc.
- 6- Crouch, e.J (1990) An approach to the automatic construction of global Thesauri. *Information Processing and management*, V 26 (5), p. 629-640.
- 7- Dubois, C.P.R. (1987) Free text vs controlled vocabulary; a reassessment. *Online Review*, v. 11, 243 - 253.
- 8- Grishman, R. (1986) *Computational* the Application of National Language Processing to Information Retrieval Tasks. *The Computer Journal* V. 35, (3), 268 - 278.
- 9- Harty, J.F. (1982) Legal Reserch using electronic techniques. In: *Lit-erature of the law - Techniques of Access (Proceedings of the 5th American Association of law Libraries Institute for law librarians, M 56 - 68)* South Hack-ensack, N.J., F.B. Rothman and Co.

- 10- Jarvelin, Kalervo (2003) Information Retrieval International Encyclopedia of Information and library Science. 2nd edited by John Feather and paul Starge, P. 293 - 295.
- 11- Johansen, T. (1987) Elements of the Non - Linguistic approach to subject relationships. *International Classification*, 14 (1): 11 - 18.
- 12- Maron, M.E. (1988) Probabilistic design principles for conventional and full-text retrieval system. *Information Processing and Management*, 24, 249 - 255.
- 13- Perez, E. (1982) Text enhancement : controlled vocabulary vs automatic text analysis. *JASIS*, 26 (1) p. 33 -44.
- 14- Wiberley, S.E. (1983). Subject access in the Humanities and the Precision of the Humanist's Vocabulary. *Library Quarterly*, 53 (4): 420-433.